Dynamic visualization of high-dimensional data via low-dimension projections and sectioning across 2D and 3D display devices

Nicholas Spyrison
Monash University, Faculty of Information Technology

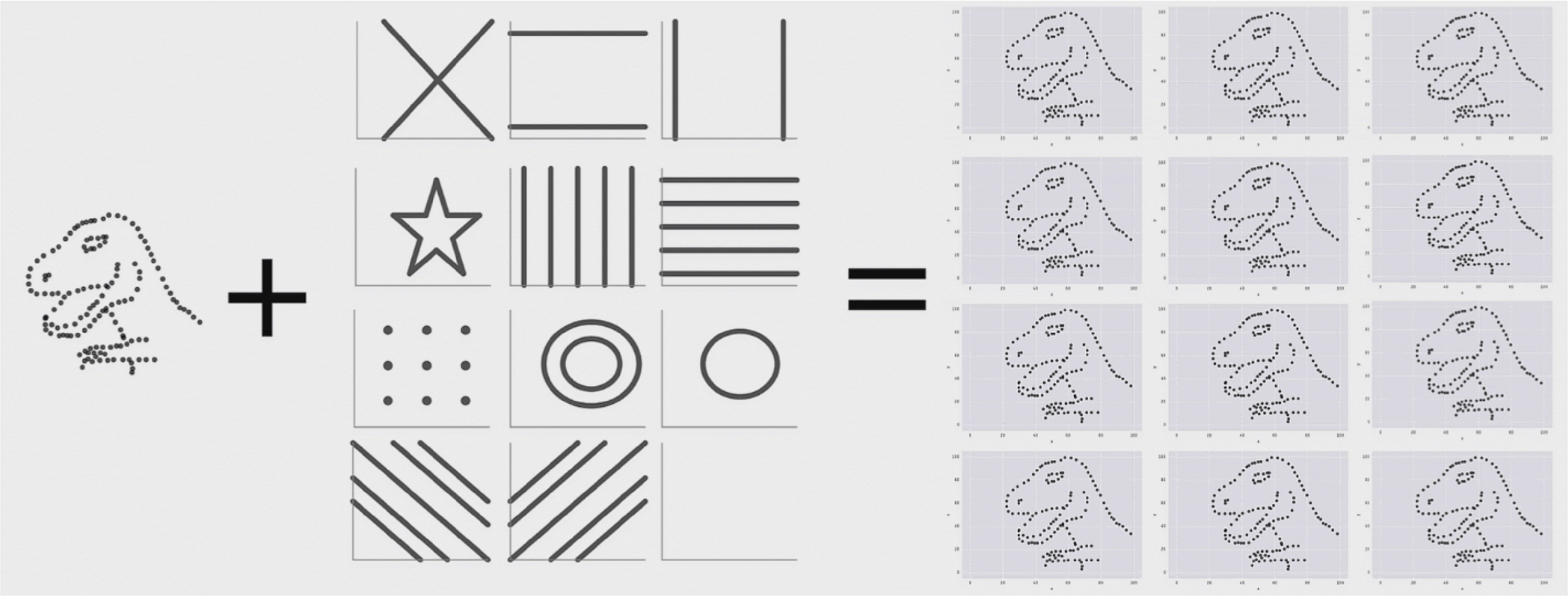
Supervisors: Prof. Kimbal Marriott,
Prof. Dianne Cook,
Prof. German Valencia

Candidature Confirmation
27 March 2019

Slides – github.com/nspyrison/confirmation_talk

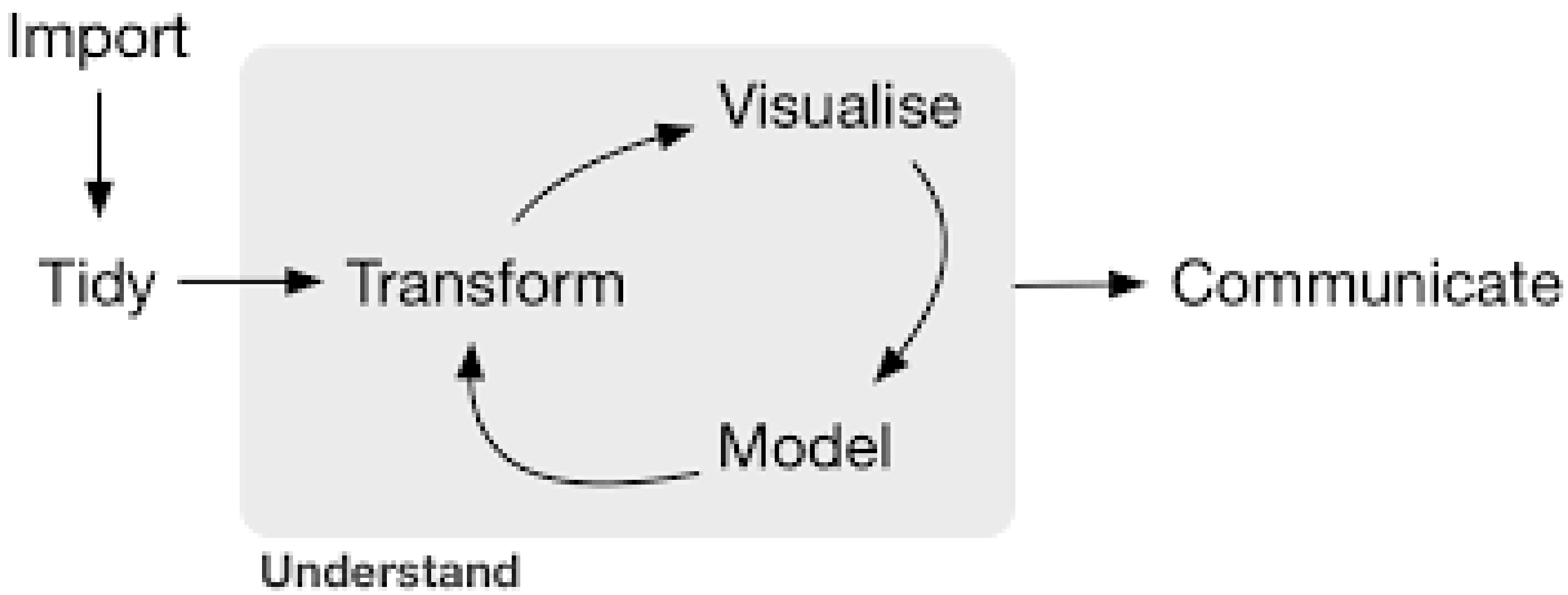
Motivation

Exploratory data analysis is important and ubiquitous, and it is important to keep visual interpretation:



Datasaurus dozen; same means, standard deviations, and correlations, *(Matejka & Fitzmaurice, 2017)*

Context: data analysis workflow



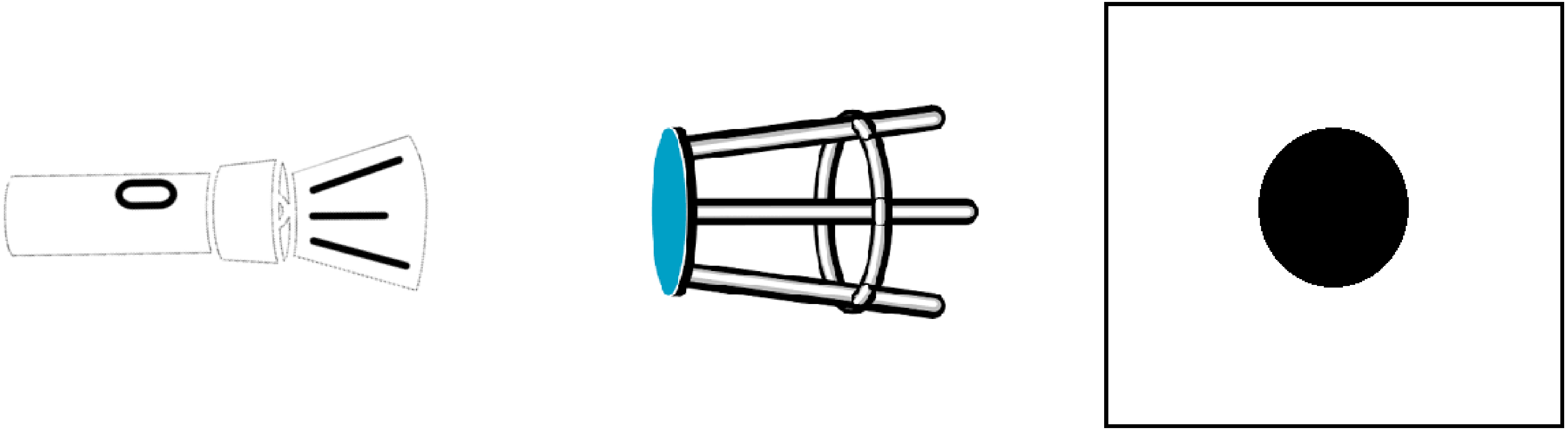
Wickham & Grolemund, 2016

- Visualization is a key aspect of data analysis workflow loop
- Reproducible from programmatic scripts (& reduced human error)
- Transparent research hosted publicly on github.

Visualizing multivariate spaces

- Visualization multivariate space is complex; dimension reduction
- Static projections do not portray all of the variation in the data
- Dynamic rotations do convey more variation and more accurate structure

Shadow puppet analogy (linear projection from 3- to 2D):

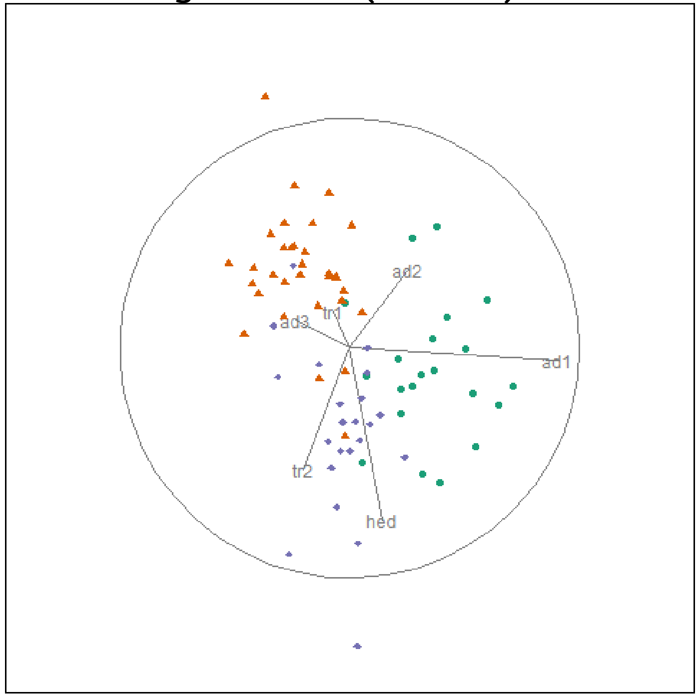


Dynamic linear projections, tours

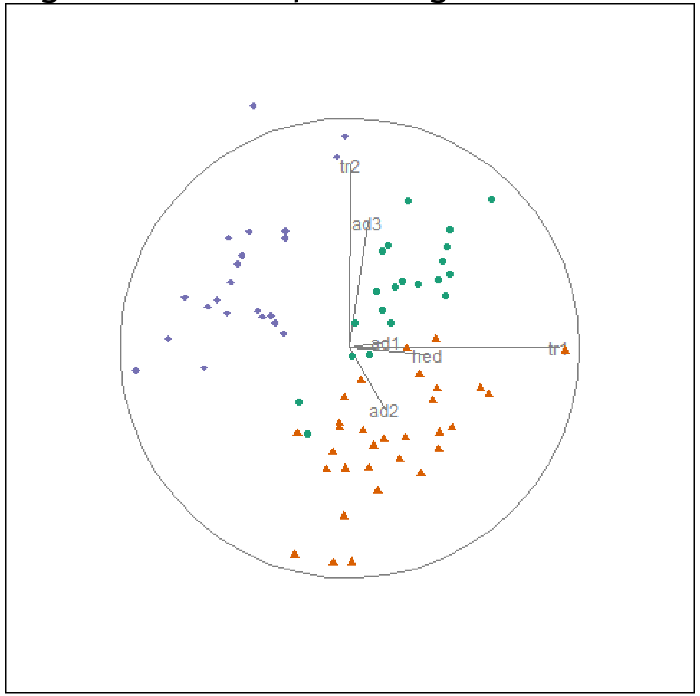
Available on CRAN, `tourr` R package, *(Wickham et al. 2011)*

- Random choice - *grand tour* random forest walk in p-space *(Asimov 1985)*
- Data-driven - *guided tour* projection pursuit, optimize an objective function on the projection *(Hurley & Buja 1990)*
- Many other geometric displays, this talk uses scatterplots

grand tour (random):

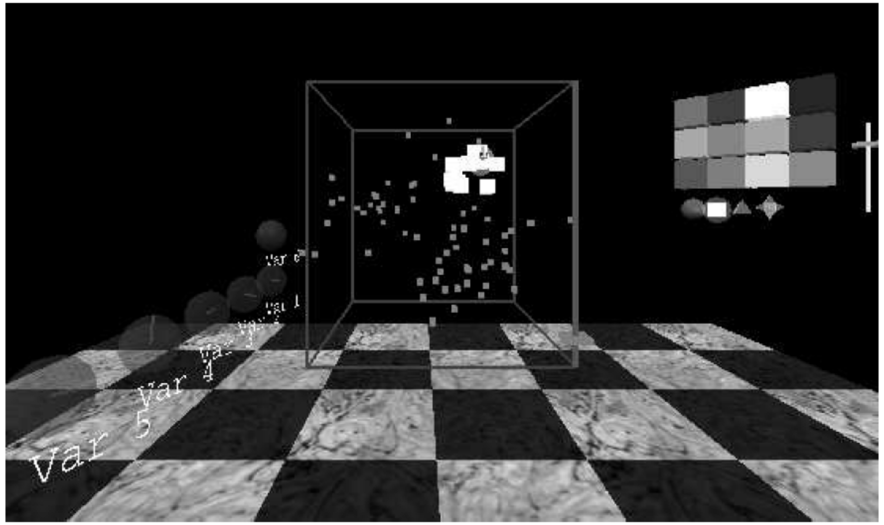


guided tour - optimizing holes index:



Visualization dimension: 3D and immersive 3D

Data visualization has lagged behind in adopting 3D and immersive technologies, despite promising finds



(*Nelson et al. 1998*)

- tours: head-tracked VR vs standard monitor
- better cluster and shape identification, slower brushing

3D visuals generally convey more information with more speed, but manipulation is slower when compared with orthogonal 2D, though 3D with 2D gives the best perception *[Lee 1986, Wickens 1994, Tory 2006]*

Embedded multivariate data in immersive 3D report improved accuracy and faster response time, but a slower manipulation speed and less comfort *[Gracia 2016, Wagner 2018, Nelson 1998]*

Modern VR equipment has improved quality, attracted wider audiences, and reduced the costs of VR, it is timely to research dynamic projections in VR

Research objectives

1) How can user-controlled steering (UCS) be generalized to work within graphic-specific environments for 2D projections?

2) Does 2D UCS tours provide benefits over alternatives?

3) How do we extend UCS to 3D?

4) Does UCS in 3D displays provide benefits over 2D displays?

RO 1) How can UCS be generalized to work within graphic-specific environments for 2D projections?

- *manual tour* user-controlled manipulation of a selected variable *(Cook & Buja 1997)*
  - Used to explore the sensitivity of the structure to the variables contributing to the projection

Work in progress. Method: Algorithm design & paper to be submitted to the R Journal:

- Algorithm generalizing for consumption by graphics environments
- R implementation via the package `spinifex`, available on github.com/nspyrison/spinifex

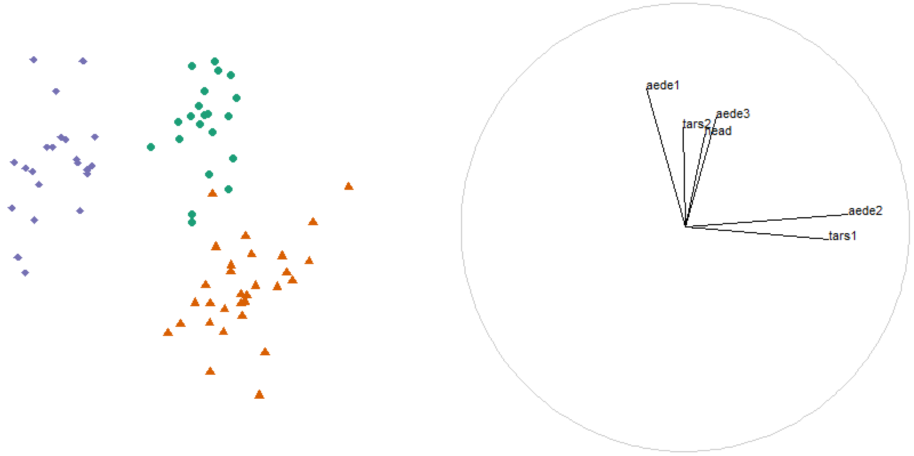
```
devtools::install_github("nspyrison/spinifex")
```

  - manual tours in R, extending the `tourr` package
  - platform to pass tours to animation-specific environments
- application to contemporary high energy physics

R01 Step 1) Choose a variable of interest

- Starting with the last projection of the previous holes-indexed guided tour:

Projected data and reference axes

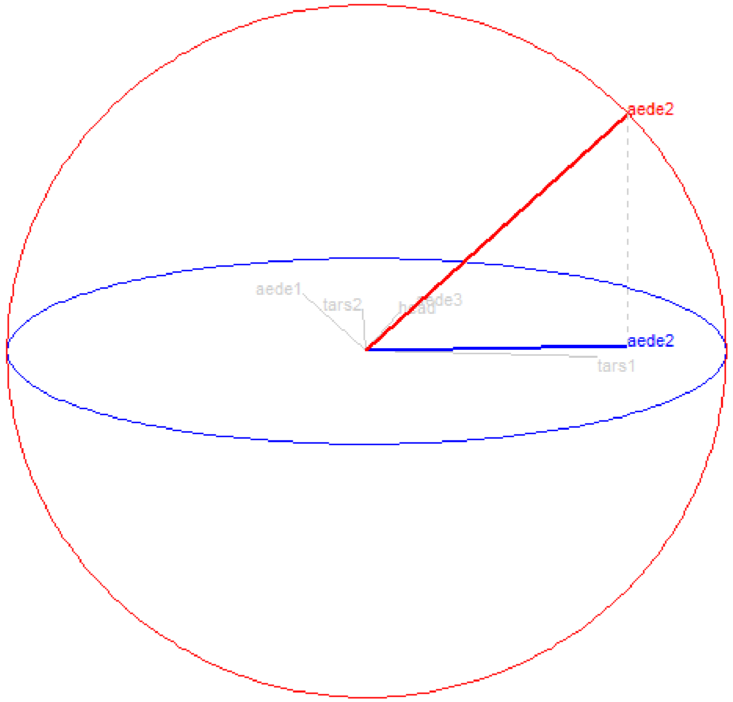


XY coefficients of each variable

	x	y
tars1	0.6423	-0.0519
tars2	-0.0100	0.4439
head	0.0886	0.4166
aede1	-0.1748	0.6180
aede2	0.7284	0.0581
aede3	0.1357	0.4914

- Choose a manipulation variable: `aede2`

RO1 Step 2) Create a manipulation space



- Orthonormalize a dimension with a full contribution to the manipulation variable
- This provides a means to rotate the basis out of the projection plane (for example, lifting paper off the table rather than being confined to the surface)
- Create a sequence of values for the 'out-of-plane' angle that will change the projection coefficients of the manipulation variable

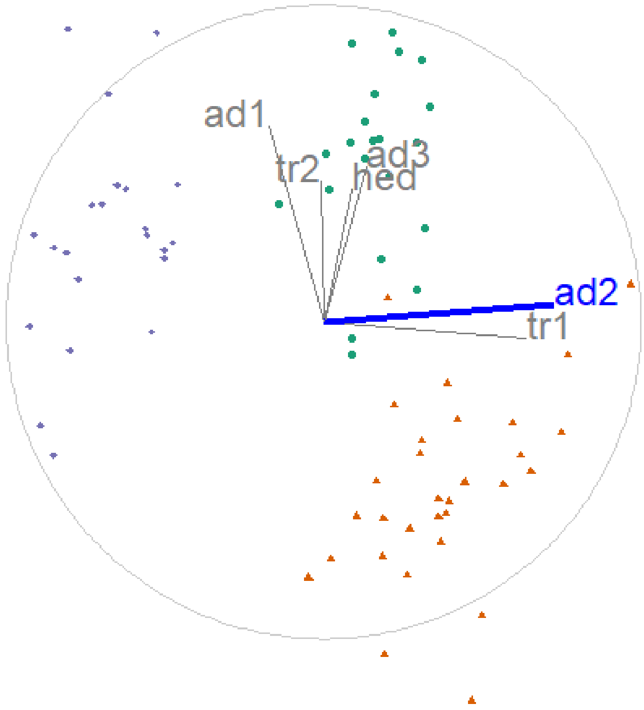
R01 Step 3) Generate the rotation

Over the sequence of angles: rotate the manipulation space for each element

- Project the data
- Plot the first two variables of the rotated basis and projection

Display as an animation

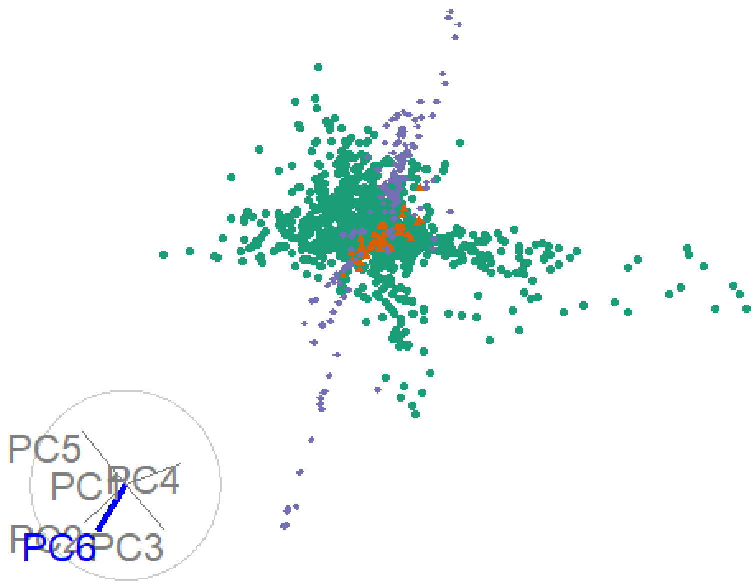
- `aede2` is important for distinguishing between the green and purple clusters



As an html widget

Application -- high energy physics

Hadronic collision experiment data, $\mathbf{X} \in \mathbb{R}^{56}$, *(Wang, et al. 2018)*, studied with guided tours *(Cook, et al. 2018)*



As an html widget, UCS on each of the 6 components

Given:

- Data summarized in 6 principal components, ~48% of the variation
- Starting basis from previously published figures

Conclusion, PC6 is important in explaining the structural features in the data:

- When the contribution of PC6 is full, the plane of green points extends into the line of sight
- When the contribution is zeroed, the line of purple points is approaching a head-on view

RO 2) What benefits does UCS provide over alternatives?

Future work, method: performance comparison assessed across benchmark datasets

- Principal Component Analysis (PCA)
  - A linear transformation that produces linear combinations of the variables in descending order of variation explained
- Multi-Dimensional Scaling (MDS)
  - Non-linear dimension reduction that compares the pairwise distance between observations
- T-distributed Stochastic Neighbor Embedding (tSNE)
  - Non-linear transformation preserves local proximity and reduces relative entropy
- User-controlled steering (USC), manual tour
  - Dynamic linear projections controlling the contribution of a selected variable

Measures: variation, variable transparency, clustering, structure

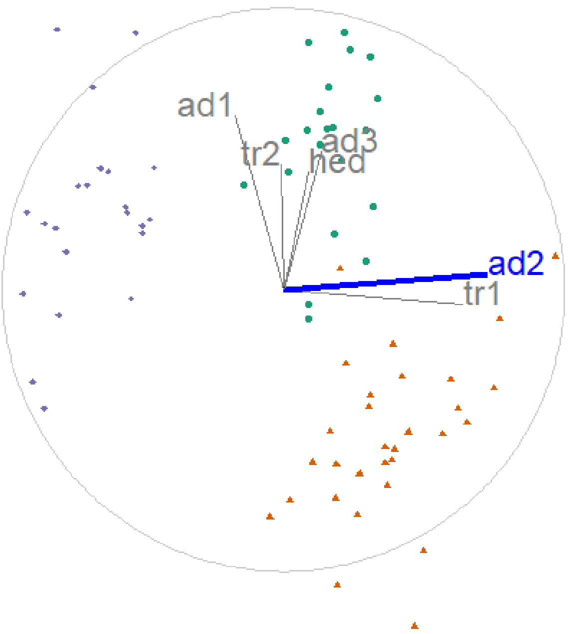
Design space: data sets, techniques, and measures of comparison

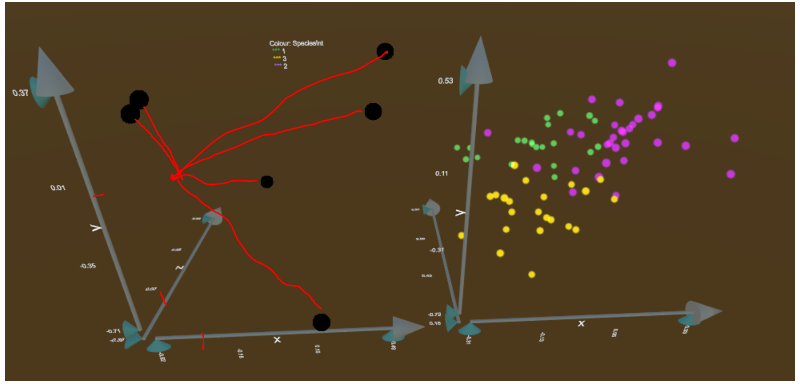
RO 3) How do we extend UCS to 3D?

Future work, method: algorithm design

Extend the UCS algorithm to 3D projections and integrate with Immersive Analytics Tool Kit, `IATK`, *(Cordeil 2019)* for a common user interface across display devices.

- Extend manipulation spaces to 4D, for variable manipulation on 3D scatterplots
- multi-dimensional function surfaces

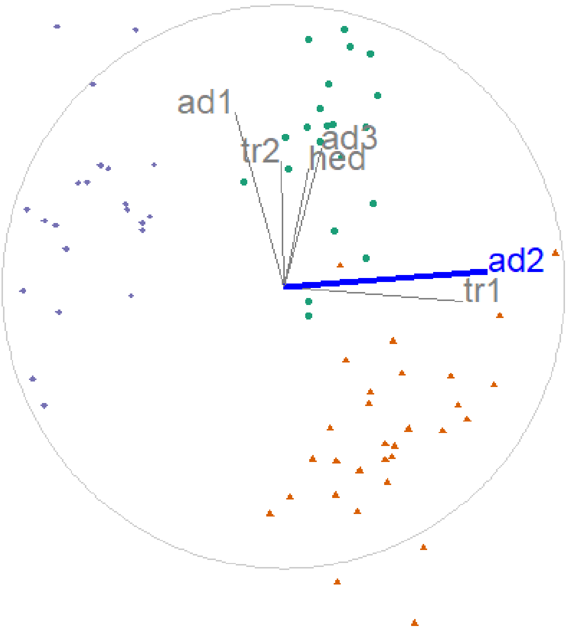


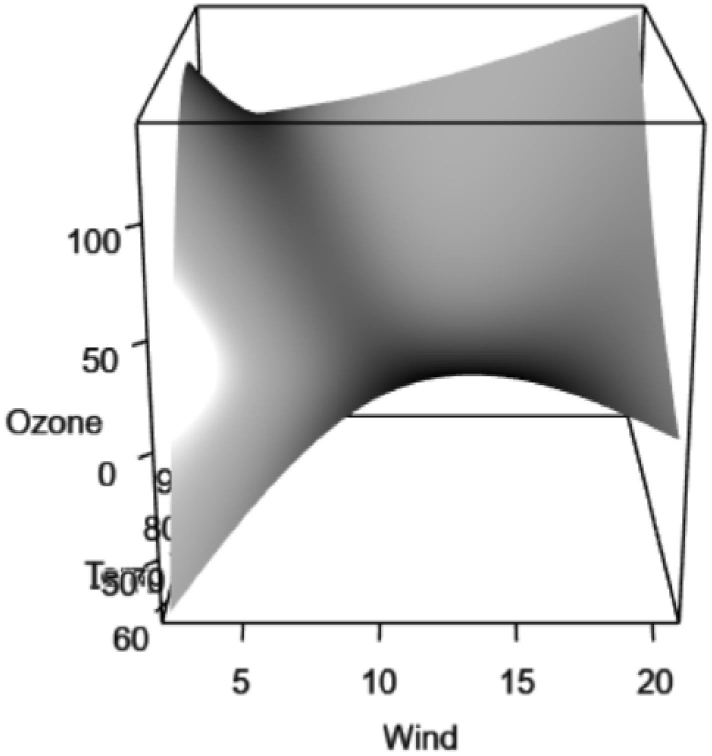


RO 3) How do we extend UCS to 3D? (& function surfaces)

Future work, method: algorithm design

- multi-dimensional function surfaces





RO 4) Does UCS in 3D displays provide benefits over 2D displays?

Future work, method: usability study

lineup design *(Hofmann et al. 2012)*

- Visual variant of statistical p-test
- Pick the real data against data generated from the null hypothesis
- Quantitative comparison across display type

Design space:

- Display type: 2D monitor, 3D monitor, head-mount, physical immersion
- Task type: structure, UCS, clustering, dimensionality
- Measures: accuracy, speed, confidence, preference, demographic information, VR and data visualization experience

Contributions

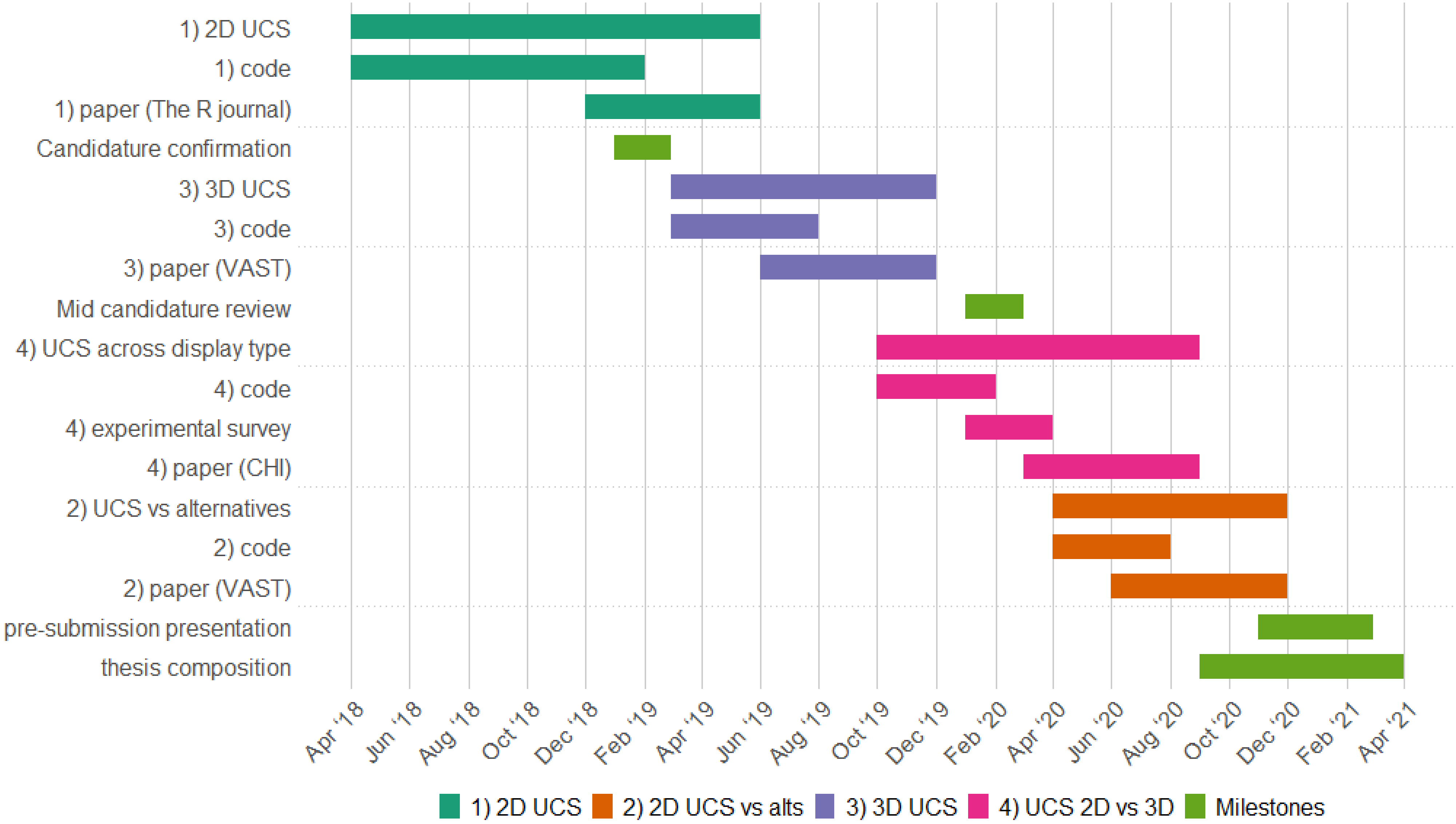
1) A modified UCS algorithm and new implementation applied to contemporary high energy physics and astrophysics applications in 2D animation frameworks.

2) A performance comparison of static and interactive UCS projection techniques assessed on benchmark data sets from the recent literature.

3) A new algorithm for UCS in 3D. With new applications to multi-dimensional function visualization in 3D.

4) Quantitative understanding of the relative benefits of UCS across 2- and 3D display devices.

Research timeline



Thanks!

Slides created in R using rmarkdown and xaringan

Slides -- github.com/nspyrison/confirmation_talk

Questions?

R Core Team, 2018

Xie et al. 2018

Xie, 2018

References (1/3)

In order of appearance:

Matejka, J., & Fitzmaurice, G. (2017). Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17 (pp. 1290–1294). Denver, Colorado, USA: ACM Press. https://doi.org/10.1145/3025453.3025912

Wickham, H., & Grolemund, G. (2016). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media, Inc.

Wickham, H., Cook, D., Hofmann, H., & Buja, A. (2011). tourr : An R Package for Exploring Multivariate Data with Projections. Journal of Statistical Software, 40(2). https://doi.org/10.18637/jss.v040.i02

Asimov, D. (1985). The Grand Tour: a Tool for Viewing Multidimensional Data. SIAM Journal on Scientific and Statistical Computing, 6(1), 128–143.

Hurley, C., & Buja, A. (1990). Analyzing High-Dimensional Data with Motion Graphics. SIAM Journal on Scientific and Statistical Computing, 11(6), 1193–1211. https://doi.org/10.1137/0911068

Nelson, L., Cook, D., & Cruz-Neira, C. (1998). XGobi vs the C2: Results of an Experiment Comparing Data Visualization in a 3-D Immersive Virtual Reality Environment with a 2-D Workstation Display. Computational Statistics, 14(1), 39–52.

References (2/3)

Lee, J. M., MacLachlan, J., & Wallace, W. A. (1986). The Effects of 3D Imagery on Managerial Data Interpretation. MIS Quarterly, 257–269.

Wickens, C. D., Merwin, D. H., & Lin, E. L. (1994). Implications of Graphics Enhancements for the Visualization of Scientific Data: Dimensional Integrality, Stereopsis, Motion, and Mesh. Human Factors, 36(1), 44–61.

Tory, M., Kirkpatrick, A. E., Atkins, M. S., & Moller, T. (2006). Visualization Task Performance with 2D, 3D, and Combination Displays. IEEE Transactions on Visualization and Computer Graphics, 12(1), 2–13.

Gracia, A., González, S., Robles, V., Menasalvas, E., & von Landesberger, T. (2016). New Insights into the Suitability of the Third Dimension for Visualizing Multivariate/Multidimensional Data: A Study Based on Loss of Quality Quantification. Information Visualization, 15(1), 3–30. https://doi.org/10.1177/1473871614556393

Wagner Filho, J., Rey, M., Freitas, C., & Nedel, L. (2018). Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-Reduced Data Scatterplots.

Cook, D., & Buja, A. (1997). Manual Controls for High-Dimensional Data Projections. Journal of Computational and Graphical Statistics, 6(4), 464–480. https://doi.org/10.2307/1390747

References (3/3)

Wang, B.-T., Hobbs, T. J., Doyle, S., Gao, J., Hou, T.-J., Nadolsky, P. M., & Olness, F. I. (2018). Visualizing the sensitivity of hadronic experiments to nucleon structure. ArXiv Preprint ArXiv:1803.02777.

Cook, D., Laa, U., & Valencia, G. (2018). Dynamical projections for the visualization of PDFSense data. Eur. Phys. J. C, 78(9), 742.

Cordeil, M. (2019). Immersive Analytics Toolkit (Version IATK 1.0 (Mala) Unity 2017). Retrieved from https://github.com/MaximeCordeil/IATK (Original work published 2017)

Hofmann, H., Follett, L., Majumder, M., & Cook, D. (2012). Graphical tests for power comparison of competing designs. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2441–2448.

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). R Markdown: The Definitive Guide. Boca Raton, Florida: Chapman and Hall/CRC. Retrieved from https://bookdown.org/yihui/rmarkdown

Xie, Y. (2018). xaringan: Presentation Ninja. Retrieved from https://CRAN.R-project.org/package=xaringan